The Duplication of the *Hox* Gene Clusters in Teleost Fishes

Sonja J. Prohaska [a] and Peter F. Stadler [a,b]

[a] *Lehrstuhl für Bioinformatik am Institut für Informatik und Interdisziplinäres Zentrum für Bioinformatik, Universität Leipzig, Kreuzstraße 7b, D-04103 Leipzig, Germany.*
{sonja,studla}@bioinf.uni-leipzig.de

[b] *Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*

Abstract

Higher teleost fishes, including zebrafish and fugu, have duplicated their *Hox* genes relative to the gene inventory of other gnathostome lineages. The most widely accepted theory contends that the duplicate *Hox* clusters orginated synchronously during a single genome duplication event in the early history of ray-finned fishes. In this contribution we collect and re-evaluate all publicly available sequence information. In particular, we show that the short *Hox* gene fragments from published PCR surveys of the killifish *Fundulus heteroclitus*, the medaka *Oryzias latipes* and the goldfish *Carassius auratus* can used to determine with little ambiguity not only their paralog group but also their membership in a particular cluster. Together with a survey of the genomic sequence data from the pufferfish *Tetraodon nigroviridis* we show that at least percomorpha, and possibly all eutelosts, share a system of seven orthologous *Hox* gene clusters, while at least the *HoxC* and *HoxD* clusters in ostariophysian (zebrafish) lineage might have arisen independently. There is little doubt about the orthology of the two teleost duplicates of the *HoxA* and *HoxB* clusters. A careful analysis of both the coding sequence of *Hox* genes and of conserved noncoding sequences provides additional support for the "duplication early" hypothesis that the *Hox* clusters in teleosts are derived by subsequent gene loss from an eight-cluster situation, although the data remain ambiguous in particular for the *HoxC* clusters. Assuming the "duplication early" hypothesis we use the new evidence on the *Hox* gene complements to determine the phylogenetic positions of gene-loss events in the wake of the cluster duplication. Surprisingly, we find that the resolution of redundancy seems to be a slow process that ist still-ongoing. A few suggestion on which additional sequence data would be most informative for resolving the history of the teleostean *Hox* genes are discussed.

Key words: *Hox* cluster, genome duplication, teleost fish, killifish, *Tetraodon nigroviridis*.

Supplemental material is available at
`http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/04-006/`.

Manuscript *13 February 2004*

1 Introduction

Hox genes code for transcription factors homologous to the genes of the Drosophila homeotic gene clusters [36, 58]. They are involved in the development of vertebrate body plan characters [59] and are one of the best-studied gene families, see e.g. [19, 35, 54, 53]. While their role in animal development is well established, their role in evolution is less well understood, see e.g. [12, 69]. A particularly intriguing problem is the role of *Hox* cluster duplications in vertebrate evolution. All invertebrates examined today have a single cluster, including the sister taxon of vertebrates, the cephalochordates, e.g. *Branchiostoma floridae* [17]. While the cluster is tightly linked in most cases, it has desintegrated in some species, including model organism such as the nematode *Caenorhabditis elegans* and the tunicate *Ciona intestinalis*.

In contrast, the ancestral *Hox* cluster was duplicated repeatedly in all extant vertebrate lineages: The common ancestor of all recent gnathostomes (sharks, bony fish, and tetrapods) had four clusters homologous to the mammalian ones [18, 49]. The four cluster situation is retained in the sarcopterygian lineage (data are available for a number of mammalia, *Xenopus tropicalis*, and both known coelacanth species [28, 4], in basal actinopterygians (bichir *Polypterus senegalus* [9]), and (presumably) also in condrichtya (horn shark, *Heterodontus francisci* [27]). Higher ray-finned fishes, however, have 6 or 7 *Hox* clusters that arose by means of duplication from the ancestral gnathostome clusters [3]. The two agnathan lineages, lampreys and hagfish, also exhibit multiple *Hox* clusters which, however, apparently arose through duplication events independent of those leading to the gnathostome clusters [24, 15, 16, 63].

Since Ohno's book on the role of gene duplication in evolution, the idea that gene and genome duplication played a major role in the origin of vertebrates has grown in support [46]. It is now clear that in fact vertebrates tend to have more copies of genes that have homologs in invertebrates and that there is also extensive variation in gene number among different clades of vertebrates [39]. Whether the duplicates of genes in vertebrates (in comparison to invertebrates) and in higher teleosts (in comparison to sarcopterygians) have arisen by means of genome duplication(s) and subsequent massiv gene loss, or whether a large number of local duplication occured has been the subject of debate. We refer to [3, 32, 34, 30, 47, 70, 71] for the genome-duplication point of view and to [51, 20] for the local duplication viewpoint. In both scenarios it is undisputed that each of the Hox clusters was duplicated as a unit.

The understanding of the evolutionary history of the actinopterygian *Hox* clusters suffers from two biasses in the available data and analyses. (1) Relatively few data (see the following section for references) are available for higher teleost species other than the two model organisms zebrafish (*Danio*

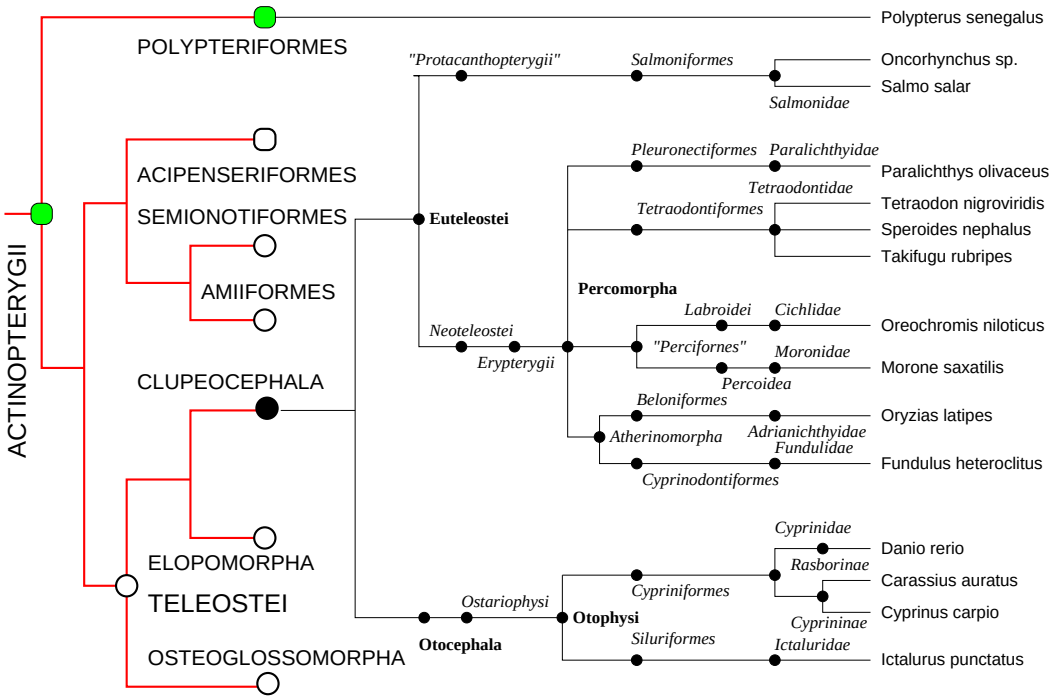

Fig. 1. Phylogenetic relationships among actinopterygian species for which *Hox* sequences are known. The tree is adapted from a recent extensive survey of mitochondrial genomes [22, 42, 60, 23]. In particular, protacanthopterygii [25] and perciformes [42] seem to be polyphyletic; see also [7] about ambiguities in teleost phylogeny. All investigated clupeocephala have duplicated *Hox* clusters (see text for references), while the bichir *Polypterus senegalus* did not share the duplication event(s). [9].

rerio) and fugu (*Takifugu rubripes*) and two of its close relatives. In particulary, very little information is available for basal groups, Fig. 1. (2) Detailed studies of *Hox* cluster evolution mostly focus on the *HoxA* cluster only, see e.g. [8, 9, 56, 61].

In this contribution we collect all publicly available sequence information on *Hox* genes of higher teleosts in order to reconstruct the history of gene losses after the teleost-specific duplication of the *Hox* clusters: (1) we provide a re-analysis of published PCR surveys that extends our knowledge about the *Hox* gene complement to additional species, (2) we report on *Hox* cluster genomics of the pufferfish *Tetraodon nigroviridis*, and (3) we re-investigate the duplication history of the teleost clusters using evidence from both coding and non-coding sequence information.

2 *Hox* Gene Inventories

2.1 Available Data

The best studied teleost species is the zebrafish *Danio rerio* whose *Hox* cluster structure is know in detail [3]: there are 49 *Hox* genes in 7 different clusters located at different chromosomes. These clusters, designated Aa, Ab, Ba, Bb, Ca, Cb, and Da, are homologous to the four mammalian Hox clusters and arose by means of one (or more) duplication events from a four cluster situation.

Recently, a thorough study on the *Hox* clusters of two pufferfishes, *Spheroides nephalus* and *Takifugu rubripes* was published [4]. Both species have a very similar structure of their *Hox* clusters as one would expect from their close phylogenetic relationship. Again there are 7 clusters at different genomic locations; in contrast to the zebrafish, however, these clusters are of the types Aa, Ab, Ba, Bb, Ca, Da, and Db.

A genome project for a third pufferfish species *Tetraodon nigroviridis* is under way [52]. Below we report on a computational survey of these publicly available data.

The third group of teleosts with extensive information on its *Hox* clusters is represented by the medaka fish *Oryzias latipes*. A genetic map containing 22 *Hox* genes shows that there are (at least) 7 *Hox clusters* located in different chromosomes, each of which is tightly linked [45]. A PCR survey [29] resulted in fragments of at least 27 distinct *Hox* genes. Recently H. Hori [1] reported that the medaka has at least 7 clusters containing a total of 46 genes organized in a way that closely resembles *Takifugu rubripes*. The details have not been published yet.

Much less is known on the *Hox* clusters of other teleost species. Fragments of the genomic sequences of the Aa and the Ba clusters are available for the striped bass (*Morone saxatilis*) [61, 57] and for the Aa cluster of tilapia (*Oreochromis niloticus*) [56]. Evidence for *HoxA2a* and *HoxA3a* genes in tilapia is reported in [34], the corresponding sequences are not available in public databases.

Systematic PCR surveys for homeobox genes were conducted for the goldfish (*Carassius auratus*) [31], the striped bass [48], the zebrafish [40], and the killifish *Fundulus heteroclitus* [41]. This technique results in short fragments (about 70-80nt) with highly conserved amino-acid sequences that in most cases can be assigned to one of the 13 vertebrate paralog groups. In the absence of

[1] http://neco.biology.kyushu-u.ac.jp/∼qshinka/poster/P-157.html

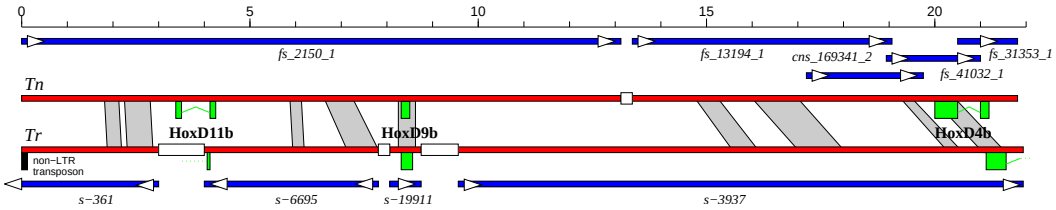

Fig. 2. Reconstruction of the *HoxDb* clusters of the two pufferfishes from the available draft assemblies. Gray areas indicate `blast` hits with $E < 10^{-20}$ and a length of at least 200nt. The total length of blast hits between the reconstructed Takifugu and Tetraodon sequences ($n = 21957$ and $n = 21808$, resp.), is 20.8% at $E = 10^{-40}$ and 29.8% at $E = 10^{-10}$.

sequences from closely related species, however, the individual cluster to which the fragments belong could not be determined. Below we revisit these sequence data and show that they can be used to determine *Hox* gene complement of the species for which no genomic information is available. A search for *Hox* genes was performed for the Atlantic Salmon *Salmo salar* resulting in the cloning of a few *Hox* genes [14].

The *HoxD10* genes of a variety of close relatives of the zebrafish were studied in [73]. Furthermore, one or the other *Hox* gene has been cloned and sequenced in the context of various studies focussing on other issues, e.g. Common carp (*Cyprinus carpio*) [65], Trout (*Oncorhynchus sp.*) [43], a flounder *Paralichthys olivaceus* [66, 67]. Finally, fragments of *Hox* genes of the following teleosts were found in Genbank: *Ictalurus punctatus*, *Salvelinus alpinus*.

A PCR survey of the stickleback *Gasterosteus aculeatus* is reported in [1]; 10 distinct *Hox* genes have been found. Unfortunately, the nucleic acid sequences are not published and no attempt has been made to distinguish between the first order paralogs.

2.2 The Third Pufferfish: Tetraodon nigroviridis

A genome sequencing project for the pufferfish *Tetraodon nigroviridis* is currently in progress by Genoscope and the Whitehead Institute for Genomic Research, see e.g. [52]. In this study we searched the draft genome assembly (version 6, release date 06 May 2002) [2].

Much of the available genomic sequence of *Tetraodon nigroviridis* is available in contigs with a length of only a few kb. The sequencing of the fugu, on the other hand, has already progressed further so that large assembled scaffolds are available. The sequences of these two species are so similar that homologous *Hox* genes can be identified unambiguously.

[2] http://www.genoscope.cns.fr/externe/tetraodon/

We find homologs of all fugu *Hox* genes with the following exceptions: (1) The version 3.0 assembly of the fugu does not seem to contain a *HoxD11a* gene, while the corresponding gene is present in the tetraodon database. (2) We were not able to identify a *HoxC1a* sequence in the tetraodon data. Apart from these two differences, which we believe are due to incomplete data, all three pufferfish species, *Tetraodon nigroviridis*, *Takifugu rubripes*, and *Spheroides nephalus* [4], appear to have the same *Hox* gene complement with the exception of an intact *HoxB7a* in *Spheroides nephalus* that has become a pseudogene in the other two pufferfish species. Evidence for a third *HoxA* cluster in *Takifugu rubripes* was reported in [4] based on a a survey of the version 2.0 assemby of the fugu genome. We were not able to find hints for such a cluster in either the version 3.0 assembly of the fugu or in the tetraodon sequence data.

The close relationship of these two pufferfish species also allows us to piece together large regions based on the assumption that the organization is the same in both species at least in regions with very high sequence homology. In a previous attempt to retrieve the *Hox* clusters of the fugu, for instance, we were not able to find the *HoxDb* cluster [50]. Careful comparison of tetraodon contigs with the fugu sequence allowed us to reconstruct this cluster for both species, Fig. 2.

In the electronic supplement to this contribution we provide preliminary reconstructions of all seven known *Hox* clusters for both *Takifugu rubripes* and *Tetraodon nigroviridis*. In some cases the assembly 3.0 of the fugu genome deviates from sequences reported from independent studies, e.g. for the region around *HoxA10a*. In these cases our reconstruction deviates from the draft assembly of the genome, see supplement for details.

2.3 PCR Surveys

The short sequences from the PCR surveys were identified by the following iterative procedure. First the 81nt long homeobox sequences were extracted from all available higher teleost sequences. An unrooted tree was computed using both neighbor joining and maximum parsimony using 1000 bootstrap replicates. Computations were performed using the `phylip` package. The resolution of this tree was sufficient to identify the paralog groups 1, 2, 3, 4, 8, 9, 10, 11, 12, and 13. In all cases the assignment of a sequence to a paralog group was cross-checked by its amino acid sequence. From the middle group (paralog groups 5, 6, and 7) sequences a separate tree was computed from which all reliable subtrees that contained a known gene were extracted. For each of these subtrees the paralog group (in many cases even the individual *Hox* cluster) was identified using the known genes located within the subtree. In the next step trees were constructed for the individual paralog groups, see

Table 1
Known *Hox* genes in higher teleosts. ■ denotes genes in genomic sequences (for fugu, tetraodon and zebrafish) spanning at least parts of the cluster. • indicate that the sequence of the gene is available at genbank but not as part of a large genomic fragment. If a gene is named in the literature but the sequence is not available the gene is marked with ▲ (most of the spheroides sequences reported in [4]). □ and △ denote known pseudogenes, – indicates the known absence of a gene. Missing data are indicated by a dot. In addition, the table contains those fragments *Hox* genes from different sources (see appendix for accession numbers) whose assignment to a cluster is unambigous or at least likely (marked with ?). A question mark without a gene name indicates less likely alternative assignment(s), see text. Gene names are taken mostly from the PCR surveys of medaka (*MF*- numbers from [29]), striped bass from [48], goldfish from [31], and killifish from [41]. The killifish sequences were extracted from the printed paper as they are not contained in the databases.

		a										b									
	Hox	Sn	Tr	Tn	Ms	On	Ol	Fh	Dr	Ca	Ss	Sn	Tr	Tn	Ms	On	Ol	Fh	Dr	Ca	Ss
	13	▲	■	■	.	■	13-3	.	■	•	trout	▲	■	■	.	.	A13b	.	■	.	.
	12	–	–	–	.	–	.	.	–	–	.	–	–	–	.	.	.	.	–	.	.
	11	▲	■	■	.	■	•	19II??	■	.	.	▲	■	■	.	.	A11b?	19II??	■		
	10	▲	■	■	■	■	•	.	□	.	.	▲	■	■	.	.	10-2	17II	■	G8-1	.
	9	▲	■	■	■	■	•	9	■	.	.	▲	■	■	.	.	•	152	■	.	.
	8	–	–	–	–	–	.	.	–	.	.	–	–	–	.	.	.	.	–	.	.
A	7	△	□	□	■	■	7-5-1??	89?	–	.	12-A	–	–	–	.	.	.		–	.	.
	6	–	–	–	–	–	.	.	–	.	.	–	–	–	.	.	.	.	–	.	.
	5	▲	■	■	■	■	•	.	■	G3-1	12-B	–	–	–	.	.	.	.	–	.	.
	4	▲	■	■	■	■	•	.	■	.	.	–	–	–	.	.	.	.	–	.	.
	3	▲	■	■	.	■	•	73?	■	.	D3x	–	–	–	.	.	.	.	–	.	.
	2	▲	■	■	.	■	2-1?	109?	□	G5-1?	.	▲	■	■	.	.	•	424	■	G5-1	.
	1	▲	■	■	.	■	•	114	■	G4-1	.	–	–	–	.	.	.	.	–	.	.
	13	•	■	■	.	.	13-2	.	■	.	.	–	–	–	.	.	.	.	–	.	.
	12	–	–	–	.	.	.	.	–	.	.	–	–	–	.	.	.	.	–	.	.
	11	–	–	–	.	.	.	.	–	.	.	–	–	–	.	.	.	.	–	.	.
	10	–	–	–	.	.	.	.	•	.	.	–	–	–	.	.	.	.	–	.	.
	9	▲	■	■	.	.	•	6	■	.	.	–	–	–	.	.	.	.	–	.	.
	8	▲	■	■	.	.	8-2	?	■	.	.	–	–	–	.	.	8-1	?	■	.	.
B	7	•	□	□	.	.	.	.	■	.	.	–	–	–	.	.	.	.	–	.	.
	6	▲	■	■	.	.	.	.	■	.	.	•	■	■	.	.	•	?	■	G2-1	.
	5	▲	■	■	A7	.	•	.	■	G1-2	.	•	■	■	C4	.	•	11’	■	G1-1	.
	4	▲	■	■	.	.	•	48	■	.	.	–	–	–	.	.	.	.	–	.	.
	3	▲	■	■	•	•	.	.	■	.	6	•	■	■	.	.	.	?	–	.	.
	2	▲	■	■	•	•	A2?		■	?	.	–	–	–	.	.	.	.	–	.	.
	1	▲	■	■	.	.	1-2	1	■	CcB1	.	•	■	■	.	.	•	288?	■	.	.
	13	▲	■	■	.	.	13-1	.	■	.	.	–	–	–	.	.	?	.	•	.	.
	12	▲	■	■	.	.	12-1	.	■	.	.	–	–	–	.	.	.	.	■	.	.
	11	▲	■	■	.	.	C11a	.	■	.	.	–	–	–	.	.	.	.	■	.	.
	10	▲	■	■	.	.	•	2’	■	G7-1	.	–	–	–	.	.	.	.	–	.	.
	9	▲	■	■	.	.	•	27	■	.	.	–	–	–	.	.	.	.	–	.	.
	8	▲	■	■	.	.	C8a	72	■	.	.	–	–	–	.	.	.	.	–	.	.
C	7	–	–	.	.	.	.	.	–	.	.	–	–	–	.	.	.	.	–	.	.
	6	▲	■	■	A4	.	A6	473	■	.	.	–	–	–	.	.	.	.	■	.	.
	5	▲	■	■	B4	.	•	282	■	.	.	–	–	–	.	.	.	.	–	.	.
	4	▲	■	■	.	.	•	349	■	.	.	–	–	–	.	.	.	.	–	.	.
	3	△	△	△	.	.	•	67	■	.	.	–	–	–	.	.	.	.	–	.	.
	2	–	–	.	.	.	.	.	–	.	.	–	–	–	.	.	.	.	–	.	.
	1	▲	▲	.	.	.	.	66?	■	.	.	–	–	–	.	.	.	.	–	.	.
	13	–	–	–	.	.	.	.	■	.	.	–	–	.	.	.	.	.	–	.	.
	12	▲	■	■	.	.	.	.	■	.	.	–	–	.	.	.	.	.	–	.	.
	11	▲	.	■	.	.	.	111II??	■	.	.	▲	■-	■	.	.	.	100II??	–	.	.
	10	▲	■	■	.	.	10-3	4 ?	■	.	.	–	–	.	.	.	.	.	–	.	.
	9	▲	■	■	.	.	•	1	■	.	.	•	■	■	.	.	•	.	–	.	.
	8	–	–	–	.	.	.	.	.	.	.	–	–	.	.	.	.	.	–	.	.
D	7	–	–	–	.	.	.	.	.	.	.	–	–	.	.	.	.	.	–	.	.
	6	–	–	–	.	.	.	.	.	.	.	–	–	.	.	.	.	.	–	.	.
	5	–	–	–	.	.	.	.	.	.	.	–	–	.	.	.	.	.	–	.	.
	4	▲	■	■	.	.	•	119	■	.	.	•	■	■	.	.	•	.	–	.	.
	3	▲	■	■	.	.	•	147	•	G11-4	.	–	–	.	.	.	.	.	–	.	.
	2	–	–	–	.	.	.		.	.	.	–	–	.	.	.	.	.	–	.	.
	1	–	–	–	.	.	.		.	.	.	–	–	.	.	.	.	.	–	.	.
Σ		47	46	46			> 42	> 28	49												

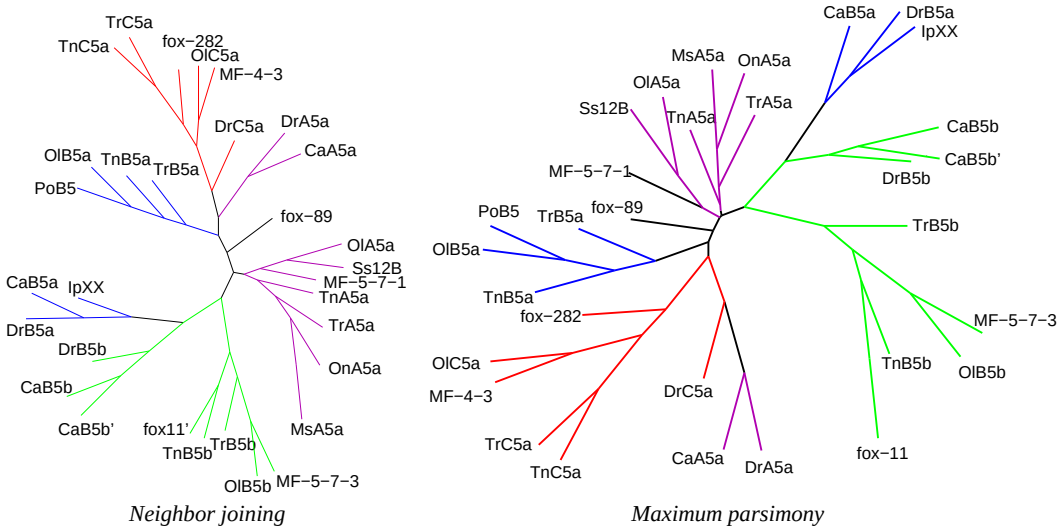

Fig. 3. Reconstructed unrooted trees for homeobox fragments (81nt) for paralog group 5. The tree reasonably identifies the genes in the euteleost and ostariophysi clades. Orthologous genes from within these two groups appear in (phylogenetically reasonable) locations in common subtrees. The resolution of the middle group genes is particularly hard. The weak phylogenetic signal in the 81nt of the homeobox fragments is insufficient to show homology of eutelost and ostariophysi sequences in all cases. MF-5-7-1 and MF-5-7-3 are most likely allelic variants for the medaka genes *HoxA5a* and *HoxB5b*. Likewise CaB5b and CaB5b' are allelic variants of the respective goldfish gene.

Fig. 3 for an example. While the phylogenetic signal is too weak in some cases to identify the orthology between euteleost and ostariophysi sequences, the identification of genes within these groups is unambiguous. The analysis is summarized in Table 1.

All goldfish genes from [31], the single carp gene, and the single catfish gene could be identified with little doubt. The same is true for most of the sequences from the PCR survey of the medaka [29]. MF-13-1 is most likely a C13a gene although we cannot rule out that it might actually be a *C13b*. We remark that a *C13a* is expected in the medaka since it is known in other euleosts, while the *C13b* gene is known only in the zebrafish. No unambigous assignment was possible for MF-5-7-1, which might be a *HoxA7a* gene.

The analysis of the killifish PCR survey [41] showed that the large majority of sequences can be reliably assigned to individual *Hox* clusters. A few clones remained ambiguous, however: fox-73 could be *Hox-B3b* or *Hox-A3a*; fox-36 could be *Hox-B8a* or *Hox-B8b*; fox-19II could be *Hox-A11a* or *Hox-A11b*. fox-100II and fox-111II both are most likely *Hox-D11* genes. This would be consistent with the discovery of a *Hox-D11b* in speroides; unfortunately the spheroides sequences are not available (yet). For a small number of clones there is some uncertainty since the sequences were placed at the most basal position of a subtree. These cases are marked with ? in Table 1.

The clone fox-89 in Fig. 3 is most likely not a *Hox5* gene but a *HoxA7a* homologous to the known PG7 genes from striped bass, tilapia, and salmon (see below). The *HoxA7* genes of *Morone saxatilis* is quite different from its sarcopterygian homologues, hence an unambiguous assignment is difficult.

2.4 Other Hox *Genes*

Flounder. The *HoxD4* gene of the flounder *Paralichthys olivaceus* is clearly homologous to the Medaka *HoxD4a* gene: In a comparison with both Medaka *HoxD4* genes and the *HoxD4* gene from *Latimeria menadoensis* [28] the tree (LmD4,OlD4b),(OlD4a,PoD4) is unambiguous. A "HoxB4" gene can be identified as *HoxB5a*.

Stickleback. The *A13a* gene of *Gasterosteus aculeatus* is an almost identical match ($E = 10^{-74}$) with the *HoxA13a* gene of the tilapia over the full length of the sequence.

Rice Field Eel. *Monopterus albus* is another percomorph fish. A PRINS study located six *Hox* clusters at six distinct chromosomes [26].

Catfish. The *B5* gene of the channel catfish *Ictalurus punctatus* is a *HoxB5a* by comparison with the first order paralogs from the zebrafish.

Salmon. The 5 homeodomain sequences that we found for the salmon are one *NK* gene (***U17652***, [62]) and 4 *Hox* genes. One of them is an EST that is identified as *HoxA3* ($E = 10^{-50}$ with the tilipia and medaka sequences). Three sequences from the survey by Fjose *et al.* [14] belong to group B3 (pS6) and two linked genes that can be unambiguously identified as *HoxA5a* (pS12-B) and *HoxA7a* (pS12-A) by combining the evidence from the aminoacid sequences, an E-value of $E = 10^{-51}$ for the comparion of ps12-B with the A5a region of the striped bass and the fact that pS12-A is located about 7.5 kb upstream of pS12-B and no homeodomain was found in between [14]. The single *Hox* gene from the trout *Oncorhynchus sp.* is *HoxA13a*.

2.5 Summary

Pufferfishes, medaka, and zebrafish have 7 *Hox* clusters that are mutually unlinked. In addition identified homologs of *Hox* genes from six different clusters for the goldfish and the killifish. All available sequence data agree with the following picture of the *Hox* cluster organization in higher teleosts:

(i) The four clusters *Hox* clusters Aa, Ab, Ba, and Bb of the clupeocephala

are true homologs and clearly have arisen by duplication from gnathostome *HoxA* and *HoxB* clusters (see below).

(ii) Euteleosts, and Percomorpha in particular, have two paralogs of the *HoxD* clusters, Da and Db, but only a single *HoxC* cluster.

(iii) Ostariophysi, in contrast, have a duplicated *HoxC* cluster but only a single *HoxD* cluster.

3 Duplication History

There is mounting evidence for a genome-wide duplication in early teleost evolution [70], which has recently been dated at about 320Myr [71], i.e., about 100Myr before divergence of zebrafish (ostariophysi) and fugu (euteleostei). Even if this is correct we cannot immediately conclude that the present *Hox* gene inventory of teleosts was determined by this event, because gene duplication is a relatively frequent, ongoing process. In the case of the *HoxA* and *HoxB* clusters there is ample evidence that they were indeed duplicated prior to the split of the euteleost and the ostariophysian lineages: Both the amino acid sequences of exon 1 of the *HoxA2*, *HoxA10*, *HoxA11*, and *HoxA13* proteins (for which both first order paralogs were retained) and the conserved non-coding DNA within the cluster show that the cluster duplication preceeded the split of the fugu and the zebrafish lineage [9]. Phylogenetic analyis of *HoxA9*, *HoxA13*, *HoxB1*, and *HoxB6* in [4] also support the duplication-first scenario. The similarity of the gene-complements of the duplicated cluster pairs, Tab. 1, as well as the retention pattern of conserved non-coding DNA sequences (so-called phylogenetic footprints [68]), Fig. 4, may serve as additional evidence.

This leaves two alternative explanations for the different cluster structure of euteleosts and ostariophysi: (1) *Duplication first*: Their common ancestor had 8 clusters of which the Db cluster was lost in ostariophysi while the Cb cluster was lost in percomorpha (the data on Salmonidae are at present insufficient to draw definite conclusions about their *Hox* gene inventory). (2) *Duplication late*: Eutelosts independently duplicated the *HoxD* cluster while in ostariophysi the *HoxC* was duplicated.

In [4] it is shown that the *HoxD4a* and *HoxD4b* sequences form distinct clusters in a neigborjoining tree; the single zebrafish *HoxD4* gene also clusters with the euteleost *HoxD4a* genes as expected in the duplication-first scenario, l.h.s. of Fig. 5. The r.h.s. of Fig. 5 displays phylogenetic networks of *HoxC9* and *HoxD9* genes computed using the *neighbor net* method [6], a generalization of the neighborjoing method for tree reconstruction implemented in the program `nnet-1.4`[3] . Phylogenetic networks computed with the *Neighbor nets* method

[3] URL: `http://www.mcb.mcgill.ca/~bryant/NeighborNet/`.

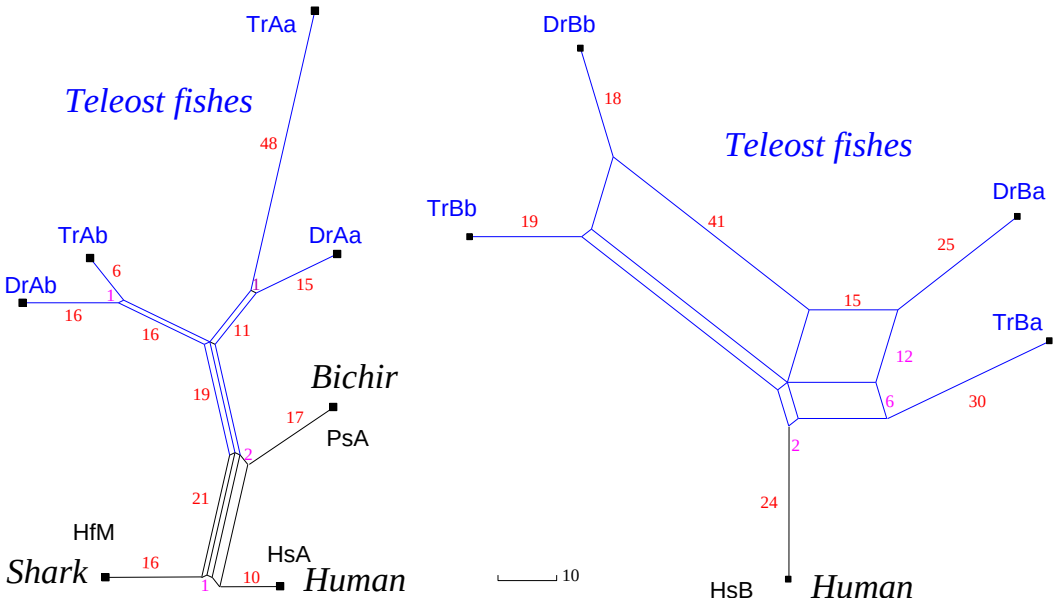

Fig. 4. Buneman graphs of the presence/absence patters of phylogenetic footprint cliques in *HoxA* and *HoxB* clusters computed using the parsimony splits methods [5] implemented in `splitstree` package [21]. Data for the *HoxA* clusters are taken from [9], data for the *HoxB* are taken from [50, Fig. 5].

reduce to the neighbor-joining [55] tree for perfectly tree-like data. On the other hand, they highlight ambiguities (alternative splits) similar to the split decomposition techniques [21] without their lack of resolution. The data are consistent with the duplication-early hypothesis for the *HoxD* cluster. They furthermore indicate a drastically increased rate of evolution in the *HoxD9b* gene prior to the common ancestor of medaka and pufferfishes. A corresponding analysis of the *HoxC* proteins remained inconclusive, although we observed tendency for the zebrafish *HoxCa* and *HoxCb* clusters to branch together.

A different line of evidence was used in [49]: The sequences of the conserved non-coding regions, i.e., the phylogenetic footprint cliques, in the *Hox* cluster also convey phylogenetic evidence that can be used independently of the coding sequences. To this end phylogenetic footprint cliques were computed separately for the *HoxA*, *HoxB*, *HoxC*, and *HoxD* clusters of mammals, frog, shark, and teleosts using the `tracker` program [50]. A list of the footprint cliques can be found in the electronic supplement. For each cluster the alignments of all individual footprint cliques are concatenated. Fig. 6 shows the phylogenetic network reconstructed using `nnet-1.4` for all four clusters.

The support for the duplication first scenario is most pronounced for the *HoxA* cluster. In the *HoxB* cluster the situation is less clearcut: The zebrafish *HoxBb* cluster branches with its putative pufferfish orthologs as expected. The position the zebrafish *HoxBa* cluster, however, is not informative. The phylogenetic net for the *HoxD* cluster is close to a "noisy star graph" with a

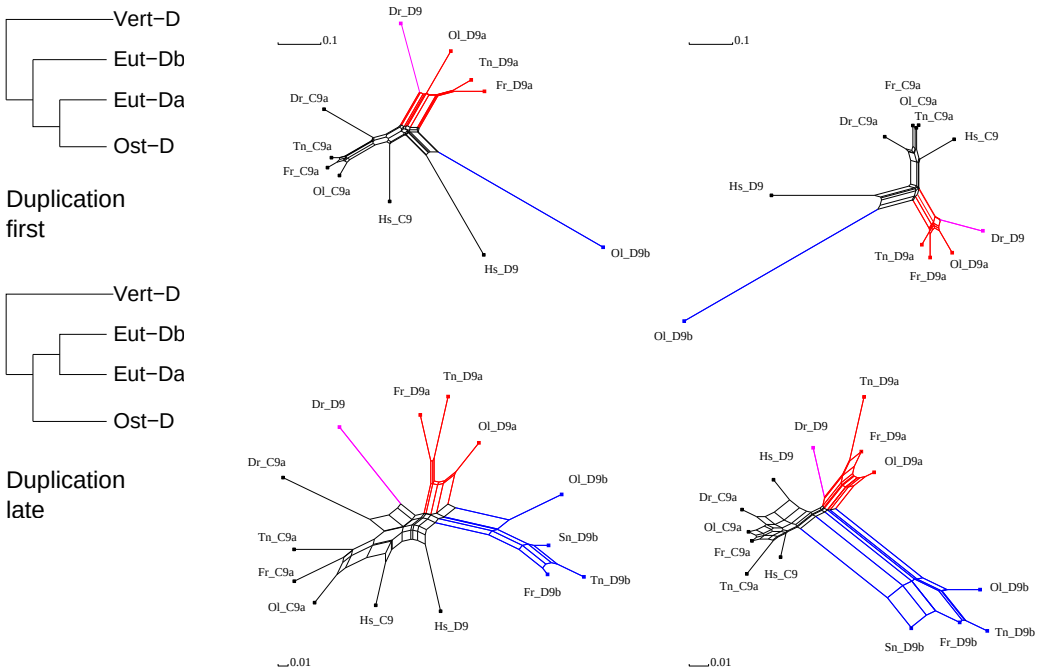

Fig. 5. Left: Scenaria for the duplication of the *HoxD* cluster in heigher teleost fishes: (1) "Duplication first" assumes the duplication of the ancestral cluster before the split of Euteleosts and the Ostariophysians. (2) "Duplication late" assumes a lineage-specific duplication for the euteleosts. Right: Phylogenetic nets of the *HoxC9* and *HoxD9* genes separatedly computed for exon 1 (above) and exon 2 (below) for all codon positions (left) and condon positions 1 and 2 only (right).

very poorly resolved interior. It is consistent with grouping the zebrafish *HoxD* cluster with the pufferfish *HoxDa* clusters, the support is very weak, however. In contrast, the data for the *HoxC* favour the duplication late scenario. In all four cases we observer that the interior branches separating the divergence of zebrafish and pufferfish lineages from the cluster duplication are very short or even not significant at all.

The hypothesis of independent, smaller-scale duplication events that took place at different times would be supported by differences of the average distance between first order paralogs of the four gnathostome *Hox* clusters. Our data indicate virtually no difference between *HoxA* and *HoxB*. The *HoxC* cluster does not significantly deviate from the over-all mean, Fig. 7. A barely significant deviation (about 3.52σ) of the average distance *HoxD* paralogs from the over-all mean value might be due to adaptive evolution in the *HoxDb* cluster of pufferfishes suggested by the long branch in Fig. 5.

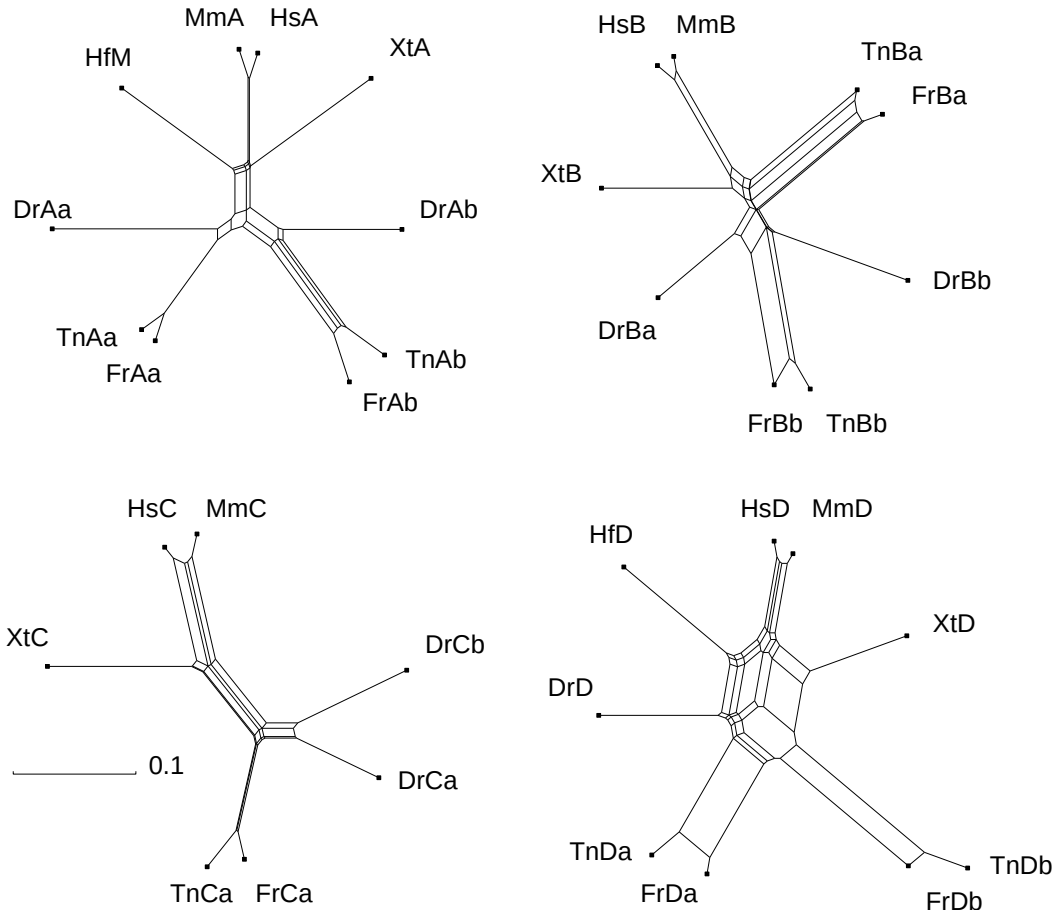

Fig. 6. Phylogenetic networks reconstructed from the sequences of the concatenated phylogenetic footprint cliques. For *HoxA* we find strong support for the "duplication first" scenario, the *HoxB* data are at least consistent with this hypothesis. In contrast, the two paralog *HoxC* clusters branch together. The network for *HoxD* is close to a noisy star; it places the zebrafish *HoxD* cluster next to the fugu and tetraodon *HoxDa* clusters but there is barely a significant split separating these three clusters from the rest.

4 Discussion

Data from all available sequences strongly support the 7 cluster situation *Aa*, *Ab*, *Ba*, *Bb*, *C*, *Da*, *Db* for the percomorpha, while ostariophysi have two *HoxC* clusters, *Ca* and *Cb*, but only a single *HoxD* cluster. The most plausible scenario given the data is a synchronous origin of all first-order paralog *Hox* clusters through a single genome duplication early in teleost evolution. As in previous studies [64, 34, 4], however, we cannot provide unambiguous evidence for this theory, albeit most of our data are consistent with this view.

The analysis of *Hox* gene duplications in teleosts is non-trivial for a number of reasons: (1) There are large differences in the rate of evolution and indications for strong adaptive evolution of certain genes in particular lineages, see e.g.

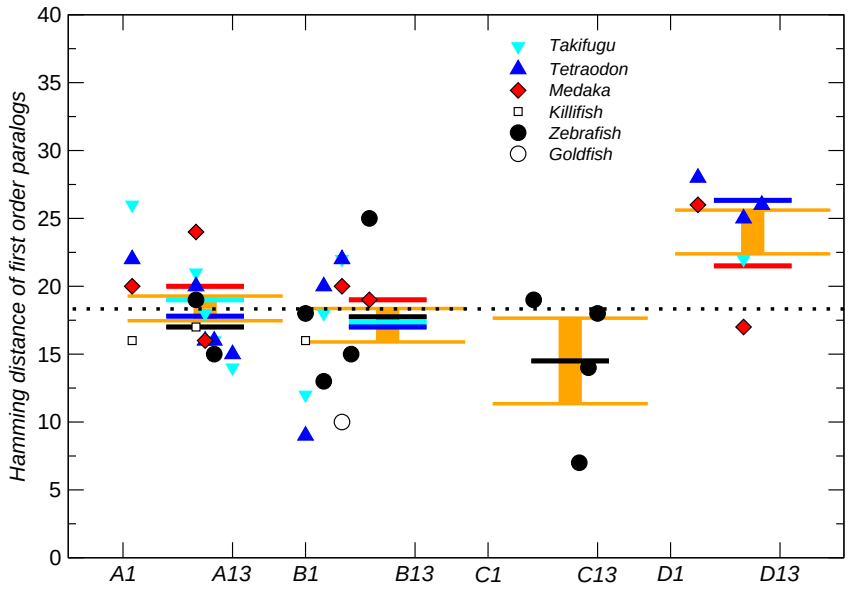

Fig. 7. The Hamming distances between first order paralogs of the 81nt fragments of the homeodomain do not show significant differences between species and between the A, B, C, and D clusters. Means for each cluster and each species are indicated by short bars, the big error bars indicated the means and standard deviations over all available first order paralog pairs from clupeocephala, see Tab. 1.

[50, 13, 33]. (2) A very uneven taxon sampling provides very few sequences that can be used to investigate the duplications of the *HoxC* and *HoxD* clusters. For instance, *HoxC* gene sequences from another ostariophysian, such as the catfish, would be very useful to confirm or disprove the possibility that the zebrafish *HoxC* cluster was duplicated late. (3) The construction of gene trees is complicated by the lack of a suitable outgroups such as paddlefish, amia, or sturgeon; data for the bichir are available for the *HoxA* cluster only.

Assuming that the duplication-first theory is correct we use the gene inventories from Tab. 1 to locate the branches along which individual *Hox* genes were lost. Surprisingly, the process of reducing the redundancy that arose through the duplication has been very slow and presumably is not yet completed [4]. This view is supported by the existence of a number of easily identifiable *Hox* pseudo-genes. In Fig. 8 we summarize the history of gene loss in the wake of the genome duplication. The loss of *HoxB7a* in some but not all pufferfishes must be very recent [4], the conversion of *HoxA7a* into a pseudo gene independently occured in zebrafish and the pufferfish lineages (and again in the bichir), [9], while the loss of *HoxB8a* occured early in part of the percomorpha lineages.

The question whether the *Hox* cluster were duplicated simultaneous could most likely be answered conclusively if a teleosts fish were discovered that has retained all 8 *Hox* clusters. The selection of a good candidate for a such a fish is closely related to the exact timing of the duplication event(s). In particular,

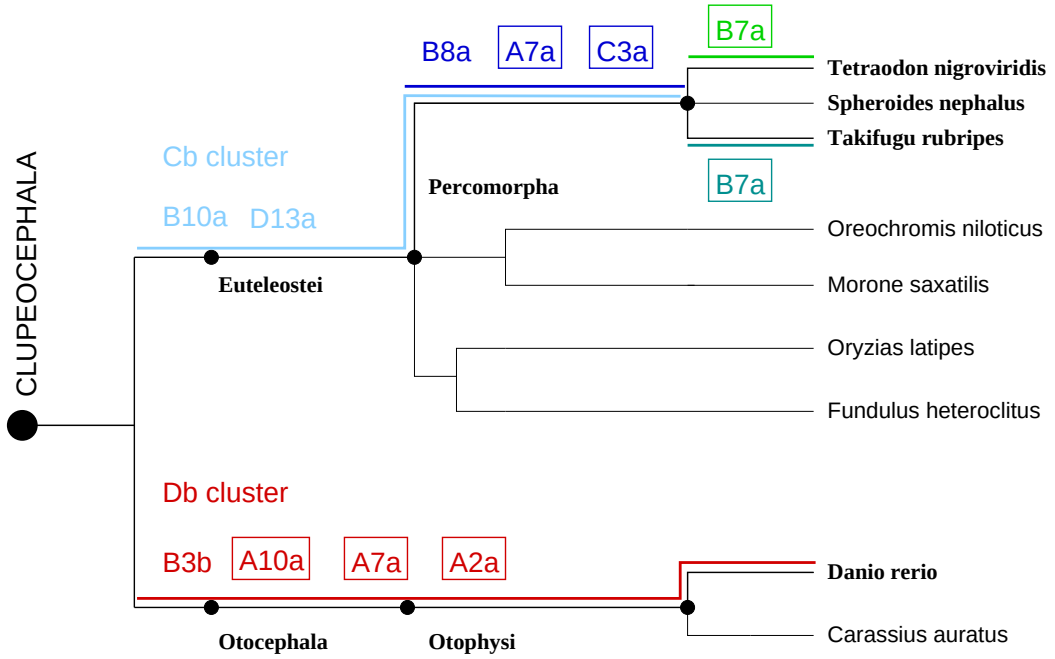

Fig. 8. Gene loss in teleost fishes. Teleost fishes for which at least a complete *Hox* gene inventory is known from genomic sequencing are shown in bold font. For the other species given only partial information is available. Colored lines indicate the maximal evolutionary time period for the loss of genes denoted by the same color. Boxed gene names represent the conversion into pseudo-genes, i.e., a loss event which is still in progress. It seems plausible to assume that the conversion of genes to pseudo-genes occurred close to the leaves of the tree, while gene loss that did not leave detectable traces occured much earlier.

it is not clear yet whether the genome duplication is causally related with the teleostean radiation.

The occurrence of duplicated ion and water transporter genes in eels [10] tentatively suggests that the duplication occurred before the common ancestor of clupeocephala and elopomorpha. Both malate dehydrogenase and triose phosphate isomerase appear in two paralog groups in higher teleosts, while the sturgeon *Acipenser brevistorum* has only a single known copy that branches outside the gene-duplication node [37, 38]. A duplicated pro-opiomelanocortin gene in paddlefish and sturgeon seems to have arisen by a chondrostean-specific duplication [11], and hence is unrelated to the teleostean genome duplication. Unpublished data mentioned in [72] also suggest a duplication before the most recent common ancestor of euteleosts and after the most recent common ancestor of the sturgeons and teleosts. These lines of still circumstantial evidence place the duplication event either immediately before or immediately after the divergence of osteoglossomorphs (for which molecular data are unfortunately very scarce) and the modern teleosts. If osteoglossomorphs and eels both should turn out to have unduplicated *Hox* clusters, sequences from both a more basal euteleost and a more basal ostariophysian fish will be required

to resolve the duplication history.

The common ancestor of all salmonids is believed to have undergone a tetraploidization event (duplication of the diploid set of chromosomes) between 25−100My [2] long after the teleostean genome duplication. This is corroborated e.g. by the existence of 4 paralog glutamine synthases in the trout, in contrast two 2 paralog genes in other teleosts and a single copy in sarcopterygians [44]. Again, the data are not clear about the exact duplication history. The recent tetraploidization can be expected to additionally complicate the analysis of salmonid sequences; as a consequence, data from a different non-percomorph eutelost would be highly desireable.

Our discussion of the available PCR fragments shows that it is feasible to determine the identity of fish *Hox* genes rather reliably already from such limited sequence information so that PCR surveys could provide very useful information despite all their problems. The resolution of the exact duplication history most probably will not require the determination of the full set of *Hox* genes from many additional species. For instance, it would be sufficient to consider *Hox11* genes in possible candidates for an 8-cluster situation since the duplication first scenario predicts that we should find 2 paralogs of both *HoxC11* and *HoxD11* along with both *HoxA11* paralogs.

Acknowledgements. Computational assistence with assembling the *Tetraodon nigroviridis Hox* clusters by Jörg Lehmann is gratefully acknowledged. This work is supported by the DFG bioinformatics initiative BIZ-6/1-2.

Appendix: Accession Numbers

Medaka (*Oryzias latipes*): ***AB026947-AB026975***, In addition to the *Hox* genes mapped in [45] there are D13x = ***AU240498***, D13y = ***AV670892***, A13x = ***AV670787***, X7x = ***AU177846***, A6 = ***AB026963***, A2 = ***AB026949***, C8A = ***AB026965***; Sequence from the PCR survey [29] can be identified with sequences from [45]: 10-3 ***AB027044*** = D10a, 10-1 ***AB027043*** = A10a, 9-2 ***AB027042*** = B9a, 9-3 ***AB027041*** = D9a, 9-4 ***AB027040*** = D9b, 9-1 ***AB027039*** = C9a, 5-7-2 ***AB027036*** = B6b, 5-7-3 ***AB027035*** = B5b, 5-7-4 ***AB027034*** = 4-3 ***AB027029*** = C5a (with 2 mutations), 5-7-5 ***AB027033*** = B5a, 5-7-6 ***AB027032*** = D4a, 4-2 ***AB027030*** = B4a, 4-1 ***AB027028*** = A4a, 3-1 ***AB027027*** = D3a, X-1 ***AB027053*** = C3a, 2-2 ***AB027026*** = A2b, 1-1 ***AB027023*** = A1a; The following database entries are identified in Tab 1: 10-2 ***AB027045***, 8-2 ***AB027038***, 8-1 ***AB027037***, 2-1 ***AB027025***, 1-2 ***AB027024***, 13-1 ***AB027047***, 13-2 ***AB027049***, 13-3 ***AB027048***, 12-1 ***AB027046***, 5-7-1 ***AB027031***, ***AU180793*** (Hox3.5), MF-C11a ***AB055740***, MF-A11b ***AB055741***, MF-A13b ***AB055742***.

Morone saxatilis: PCR survey: [48] B4 ***U09950***, C4 ***U09949***, B5 ***U09948***, A7

U09947, A6 ***U09946***, A4 ***U09945***, B3 ***U09944***, A3 ***U09943***; genomic fragments: Aa cluster ***AF089743***, Ba cluster ***AF517833***

Goldfish (*Carassius auratus*) [31]: G1-1 ***L09685***, G1-2 ***L09686***, G2-1 ***L09687***, G3-1 ***L09688***, G3-2 ***L09689*** (variant of G1-1), G5-1 ***L09690***, G5-2 ***L09691***, G7-1 ***L09693***, G8-1 ***L09694***, G11-4 ***L09697***, G4-1 ***L09698***.

Common carp (*Cyprinus carpio*) [65] CcB1x ***X91079***.

Atlantic salmon (*Salmo salar*) SsD3x AcBG933993 (EST database), pS12-A SsX7x ***M18903***, pS12-B ***M18904***, pS6 ***M18905*** [14];

Trout (*Oncorhynchus sp.*) OxA13 ***AF107229***.

Channel catfish (*Ictalurus punctatus*): ***BM424931*** B5a.

Japanese Flounder (*Paralichthys olivaceus*): PoD4 ***AB029749*** [66], PoB5 ***AB029759*** [67].

Three-spined stickleback (*Gasterosteus aculeatus*): GaA13 ***AF107228***.

Southern puffer (*Spheroides nephalus*) [4]: SnB3b ***AY303235***, SnB6B ***AY303234***, SnB13a ***AY303233***, SnD4b ***AY303232***, SnD9b ***AY303231***, SnB7a ***AY303230***.

References

[1] D.-g. Ahn and G. Greg. Expression patterns of threespine stickleback *Hox* genes and insights into the evolution of the vertebrate body axis. *Dev. Genes Evol.*, 209:482–494, 1999.

[2] F. W. Allendorf and G. H. Thorgaard. Tetraploidy and the evolution of salmonid fishes. In B. J. Turner, editor, *Evolutionary Genetics of Fishes*, pages 1–46. Plenum Press, New York, 1984.

[3] A. Amores, A. Force, Y. L. Yan, L. Joly, C. Amemiya, A. Fritz, R. K. Ho, J. Langeland, V. Prince, Y. L. Wang, M. Westerfield, M. Ekker, and J. H. Postlethwait. Zebrafish *Hox* clusters and vertebrate genome evolution. *Science*, 282:1711–1714, 1998.

[4] A. Amores, T. Suzuki, Y.-L. Yan, J. Pomeroy, A. Singer, C. Amemiya, and J. Postlethwait. Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-fin fish. *Genome Res.*, 14:1–10, 2004.

[5] H.-J. Bandelt and A. W. M. Dress. A relational approach to split decomposition. In O. Opitz, B. Lausen, and R. Klar, editors, *Information and Classification*, pages 123–131. Springer-Verlag, Berlin, 1993.

[6] D. Bryant and V. Moulton. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.*, 2004. doi: 10.1093/molbev/msh018.

[7] W.-J. Chen, C. Bonillo, and G. Lecointre. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of acanthomorpha (teleostei) with larger number of taxa. *Mol. Phylog. Evol.*, 26:262–288, 2003.

[8] C.-h. Chiu, C. Amemiya, K. Dewar, C.-B. Kim, F. H. Ruddle, and G. P. Wagner. Molecular evolution of the HoxA cluster in the three major gnathostome lineages. *Proc. Natl. Acad. Sci. USA*, 99:5492–5497, 2002.

[9] C.-H. Chiu, K. Dewar, G. P. Wagner, K. Takahashi, F. Ruddle, C. Ledje, P. Bartsch, J.-L. Scemama, E. Stellwag, C. Fried, S. J. Prohaska, P. F. Stadler, and C. T. Amemiya. Bichir *HoxA* cluster sequence reveals surprising trends in rayfinned fish genomic evolution. *Genome Res.*, 14:11–17, 2004.

[10] C. P. Cutler and G. Cramb. Molecular physiology of osmoregulation in eels and other teleosts: the role of transporter isoforms and gene duplication. *Comp. Biochem. Physiology A*, 130:551–564, 2001.

[11] P. B. Danielson, J. Alrubaian, M. Muller, J. M. Redding, and R. M. Dores. Duplication of the POMC gene in the paddlefish (*Polyodon spathula*): Analysis of γ-msh, acth, and β-endorphin regions of ray-finned fish POMC. *Gen. Comp. Endocrin.*, 132:384–390, 1999.

[12] E. Davidson. *Genomic Regulatory Systems*. Academic Press, San Diego, 2001.

[13] M. A. Fares, D. Bezemer, A. Moya, and I. Marín. Selection on coding regions determined *Hox7* genes evolution. *Mol. Biol. Evol.*, 20:2104–2112, 2003.

[14] A. Fjose, A. Molven, and H. G. Eiken. Molecular cloning and characterization of homeo-box-containing genes from atlantic salmon. *Gene*, 62:141–152, 1988.

[15] A. Force, A. Amores, and J. H. Postlethwait. Hox cluster organization in the jawless vertebrate *Petromyzon marinus*. *J. Exp. Zool. (Mol. Dev. Evol.)*, 294:30–46, 2002.

[16] C. Fried, S. J. Prohaska, and P. F. Stadler. Independent hox-cluster duplications in lampreys. *J. Exp. Zool., Mol. Dev. Evol*, 299B:18–25, 2003.

[17] J. Garcia-Fernández and P. W. Holland. Archetypal organization of the amphioxus hox gene cluster. *Nature*, 370:563–566, 1994.

[18] P. W. Holland and J. Garcia-Fernandez. Hox genes and chordate evolution. *Dev. Biol.*, 173:382–395, 1996.

[19] P. W. H. Holland, J. Garcia-Fernández, N. A. Williams, and A. Sidow. Gene duplication and the origins of vertebrate development. *Development*, (Suppl.):125–133, 1994.

[20] A. L. Hughes and R. Friedman. 2R or not 2R: testing hypotheses of genome duplication in early vertebrates. *J. Struct. Funct. Genomics*, 3:85–93, 2003.

[21] D. H. Huson. Splitstree: analyzing and visualizing evolutionary data. *Bioinformatics*, 14:68–73, 1998.

[22] J. G. Inoue, M. Miya, K. Tsukamoto, and M. Nishida. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the " ancient fish". *Mol. Phylog. Evol.*, 26:110–120, 2003.

[23] J. G. Inoue, M. Miya, K. Tsukamoto, and M. Nishida. Mitogenomic evidence for the monophyly of elopomorph fishes (telostei) and the evolutionary origin of the leptocephalus larva. *Mol. Phylog. Evol.*, 2004. doi: 10.1016/j.ympev.2003.11.009.

[24] S. Q. Irvine, J. L. Carr, W. J. Bailey, K. Kawasaki, N. Shimizu, C. T. Amemiya, and F. H. Ruddle. Genomic analysis of Hox clusters in the sea lamprey, *Petromyzon marinus*. *J. Exp. Zool. (Mol. Dev. Evol.)*, 294:47–62, 2002.

[25] N. B. Ishiguro, M. Miya, and M. Nishida. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". *Mol. Phylog. Evol.*, 27:476–488, 2003.

[26] F. Y. Ji, J. D. Liu, M. S. Yi, L. Huang, F. Zhou, and Q. X. Yu. Chromosomal localization of rice field eel *Hox* genes by PRINS. *Yi Chuan Xue Bao (Acta Genetica Sinica)*, 29:612–615, 2002. (chinese).

[27] C. B. Kim, C. Amemiya, W. Bailey, K. Kawasaki, J. Mezey, W. Miller, S. Minosima, N. Shimizu, W. G. P., and F. Ruddle. Hox cluster genomics in the horn shark, *heterodontus francisci*. *Proc. Natl. Acad. Sci. USA*, 97:1655–1660, 2000.

[28] E. G. L. Koh, K. Lam, A. Christoffels, M. V. Erdmann, S. Brenner, and B. Venkatesh. *Hox* gene clusters in the indonesian coelacanth, *Latimeria menadoensis*. *Proc. Natl. Acad. Sci. USA*, 100:1084–1088, 2003.

[29] G. Kurosawa, K. Yamada, H. Ishiguro, and H. Hori. *Hox* gene complexity in medaka fish may be similar to that in pufferfish rather than zebrafish. *Biochem. Biophys. Res. Commun.*, 260:66–70, 1999.

[30] D. Larhammar, L. G. Lundin, and H. F. The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. *Genome Res.*, 12:1910–1920, 2002.

[31] E. M. Levine and N. Schechter. Homeobox genes expressed in the retina and brain of adult goldfish. *Proc. Natl. Acad. Sci. USA*, 90:2729–2733, 1993.

[32] M. Lynch and J. S. Conery. The evolutionary fate and consequences of duplicate genes. *Science*, 290:1151–1155, 2000.

[33] V. J. Lynch, J. J. Roth, K. Takahashi, C. Dunn, D. Nonaka, G. Stopper, and G. P. Wagner. The origin of placental mammals is coincident with adaptive evolution of the developmental control genes *HoxA-11* and *HoxA-13*. 2004. submitted.

[34] E. Málaga-Trillo and A. Meyer. Genome duplications and accelerated evolution of *Hox* genes and cluster architecture in teleost fishes. *Amer. Zool.*, 41:676–686, 2001.

[35] P. Martinez and C. T. Amemiya. Genomics of the HOX gene cluster. *Comp. Biochem. Physiol. Part B*, 133:571–580, 2002.

[36] W. McGinnis and R. Krumlauf. Homeobox genes and axial patterning.

Cell, 68:283–302, 1992.

[37] T. J. S. Merrit and J. M. Quattro. Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase. *Genetics*, 159:689–697, 2001.

[38] T. J. S. Merrit and J. M. Quattro. Evolution of the vertebrate cytosolic malate dehydrogenase gene family: Duplication and divergence in actinopterygian fish. *J. Mol. Evol.*, 56:265–276, 2003.

[39] A. Meyer and M. Schartl. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.*, 11:699–704, 1999.

[40] B. Y. Misof, M. J. Blanco, and G. P. Wagner. A PCR-survey of *Hox* genes of the zebrafish: new sequences and evolutionary implications. *J. Exp. Zool.*, 274:193–206, 1996.

[41] B. Y. Misof and G. P. Wagner. Evidence for four Hox clusters in the killifish *Fundulus Heteroclitus* (teleostei). *Mol. Phylog. Evol.*, 5:309–322, 1996.

[42] M. Miya, H. Takeshima, H. Endo, N. B. Ishiguro, J. G. Inoue, T. Mukai, T. P. Satoh, M. Yamaguchi, A. Kawaguchi, K. Mabuchi, S. M. Shirai, and M. Nishida. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial dna sequences. *Mol. Phylog. Evol.*, 26:121–138, 2003.

[43] D. P. Mortlock, P. Sateesh, and J. W. Innis. Evolution of N-terminal sequences of the vertebrate HOXA13 protein. *Mamm. Genome*, 11:151–158, 2000.

[44] B. W. Murray, E. R. Busby, T. P. Mommsen, and P. A. Wright. Evolution of glutamine synthetase in vertebrates: multiple glutamine synthetase genes expressed in rainbow trout (*Oncorhynchus mykiss*). *J. Exp. Biol.*, 206:1511–1521, 2003.

[45] K. Naruse, S. Fukamachi, H. Mitani, M. Kondo, T. Matsuoka, S. Kondo, N. Hanamura, Y. Morita, K. Hasegawa, R. Nishigaki, A. Shimada, H. Wada, T. Kusakabe, N. Suzuki, M. Kinoshita, A. Kanamori, T. Terado, H. Kimura, M. Nonaka, and A. Shima. A detailed linkage map of medaka, *Oryzias latipes*: Comparative genomics and genome evolution. *Genetics*, 154:1773–1784, 2000.

[46] S. Ohno. *Evolution by Gene Duplication*. Springer Verlag, New York, 1970.

[47] G. Panopoulou, S. Hennig, D. Groth, A. Krause, A. J. Poustka, R. Herwig, M. Vingron, and H. Lehrach. New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res.*, 13:1056–1066, 2003.

[48] A. M. Pavell and E. J. Stellway. Survey of *Hox*-like genes in the teleost *Morone saxatilis*: Implications for the evolution of the *Hox* gene family. *Mar. Mol. Biol. Biotech.*, 3:149–157, 1994.

[49] S. J. Prohaska, C. Fried, C. T. Amemiya, F. H. Ruddle, G. P. Wagner, and P. F. Stadler. The shark HoxN cluster is homologous to the human

HoxD cluster. *J. Mol. Evol.*, 2004. in press.

[50] S. J. Prohaska, C. Fried, C. Flamm, G. Wagner, and P. F. Stadler. Surveying phylogenetic footprints in large gene clusters: Applications to Hox cluster duplications. *Mol. Phyl. Evol.*, 2003. in press; doi: 10.1016/j.ympev.2003.08.009.

[51] M. Robinson-Rechavi, M. O, H. Escriva, P. L. Bardet, D. Zelus, S. Hughes, and V. Laudet. Euteleost fish genomes are characterized by expansion of gene families. *Genome Res.*, 11:781–788, 2001.

[52] H. Roest Crollius, O. Jaillon, C. Dasilva, C. Ozouf-Costaz, C. Fizames, C. Fischer, L. Bouneau, A. Billault, F. Quetier, W. Saurin, A. Bernot, and J. Weissenbach. Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Research*, 10:939–949, 2000.

[53] F. H. Ruddle, J. L. Bartels, K. L. Bentley, C. Kappen, M. T. Murta, and P. J. W. Evolution of Hox genes. *Annu. Rev. Genet.*, 28:423–442, 1994.

[54] F. H. Ruddle, K. L. Bentley, M. T. Murtha, and N. Risch. Gene loss and gain in the evolution of the vertebrates. *Development*, (Supplement):155–161, 1994.

[55] N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol. Evol.*, 4:406–425, 1987.

[56] S. Santini, J. L. Boore, and A. Meyer. Evolutionary conservation of regulatory elements in vertebrate *Hox* gene clusters. *Genome Res.*, 13:1111–1122, 2003.

[57] J.-L. Scemama, M. Hunter, J. McCallum, V. Prince, and E. Stellwag. Evolutionary divergence of vertebrate *Hoxb2* expression patterns and transcriptional regulatory loci. *J. Exp. Zool., Mol. Dev. Evol*, 294:285–299, 2002.

[58] F. R. Schubert, K. Nieselt-Struwe, and P. Gruss. The antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *Proc. Natl. Acad. Sci. USA*, 90:143–147, 1993.

[59] C. S. Shashikant, M. F. Utset, T. L. Violette, S. M. Wise, P. Einat, M. Einat, J. W. Pendleton, K. Schughart, and F. H. Ruddle. Homeobox genes in mouse development. *Crit. Rev. Eukaryot. Gene Expr.*, 1:207–245, 1991.

[60] M. P. Simmons and M. Miya. Efficiently resolving the basal clades of a phylogenetic tree using bayesian and parsimony approaches: a case study using mitogenomic data from 100 higher teleost fishes. *Mol. Phylog. Evol.*, in press. doi:10.1016/j.ympev.2003.08.004.

[61] E. A. Snell, J. L. Scemama, and E. J. Stellwag. Genomic organization of the *Hoxa4-Hoxa10* region from *Morone saxatilis*: implications for *Hox* gene evolution among vertebrates. *J. Exp. Zool. (Mol. Dev. Evol.)*, 285:41–49, 1999.

[62] H. Stadler, J. C. Murray, N. J. Leysens, P. J. Goodfellow, and M. Solursh. Phylogenetic conservation and physical mapping of members of the H6 homeobox gene family. *Mamm. Genome*, 6:383–388, 1995.

[63] P. F. Stadler, C. Fried, S. J. Prohaska, W. J. Bailey, B. Y. Misof, F. H. Ruddle, and G. P. Wagner. Evidence for independent *Hox* gene duplications in the hagfish lineage: A PCR-based gene inventory of *Eptatretus stoutii*. *Mol. Phylog. Evol.*, 2003. submitted.

[64] E. J. Stellwag. Hox gene duplications in fish. *Cell Devel. Biol.*, 10:531–540, 1999.

[65] C. J. Stevens, J. Samallo, H. Schipper, H. W. Stroband, and G. te Kronnie. Expression of *Hoxb-1* during gastrulation and segmentation stages of carp (*Cyprinus carpio*). *Int. J. Dev. Biol.*, 40:463–470, 1996.

[66] T. Suzuki, I. Oohara, and T. Kurokawa. *Hoxd-4* expression during pharyngeal arch development in flounder (*Paralichthys olivaceus*) embryos and effects of retinoic acid on expression. *Zoolog Sci.*, 15:57–67, 1998.

[67] T. Suzuki, A. S. Srivastava, and T. Kurokawa. *Hoxb-5* is expressed in gill arch 5 during pharyngeal arch development of flounder *Paralichthys olivaceus* embryos. *Int. J. Dev. Biol.*, 43:357–359, 1999.

[68] D. A. Tagle, B. F. Koop, M. Goodman, J. L. Slightom, D. L. Hess, and R. T. Jones. Embryonic epsilon and gamma globin genes of a prosimian primate (galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J. Mol. Biol.*, 203:439–455, 1988.

[69] Y. Takahashi, J.-i. Hamada, K. Murakawa, M. Takada, M. Tada, I. Nogami, N. Hayashi, S. Nakamoric, M. Monden, M. Miyamoto, H. Katoh, and T. Moriuchi. Expression profiles of 39 *HOX* genes in normal human adult organs and anaplastic thyroid cancer cell lines by quantitative real-time RT-PCR system. *Exp. Cell. Res.*, 293:144–153, 2004.

[70] J. Taylor, I. Braasch, T. Frickey, A. Meyer, and Y. Van De Peer. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.*, 13:382–390, 2003.

[71] K. Vandepoele, W. De Vos, J. S. Taylor, A. Meyer, and Y. Van de Peer. Major events in the genome evolution of vertebrates: Paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl. Acad. Sci. USA*, 101:doi 10.1073/pnas.0307968100, 2004.

[72] G. P. Wagner, C. Amemiya, and F. Ruddle. Hox cluster duplication and the genetics of evolutionary novelties. *Proc. Natl. Acad. Sci.*, 100:14603–14606, 2003.

[73] R. Zardoya, E. Abouheif, and A. Meyer. Evolutionary analyses of hedgehog and *Hoxd-10* genes closely related to the zebrafish. *Proc. Natl. Acad. Sci. USA*, 93:13036–13041, 1996.